IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 11, Issue 6, November-December 2024

Impact Factor: 7.394

www.ijarety.in editor.ijarety@gmail.com

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106023

Vector Borne Diseases Outbreak Prediction

P. Anusha[1], Aleti Manisha[2], Eerla Divya Patel[3], Kalyan Gundedi[4]

Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India[1]

Student, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India[2,3,4]

ABSTRACT: Vector Borne diseases are rising challenge in India. These have become a burden for the society and prevention and control of these vector borne diseases is still a challenge for the government. Large portion of population is India is infected from this disease every year. Due to the diversity in geographical and living standard of people, it becomes difficult to control these diseases at early stages in the present system. The goal of this study is to investigate the symptoms and study the influence of clinical test parameters that belong to vector borne disease. The main aim of this study is to develop a prediction model using machine learning techniques for vector borne diseases. The prediction model has been prepared with the help of Logistic Regression. At the end, study proposes a prediction model for vector borne disease capable of diagnosing Vector Borne disease at each stage of Vector Borne Disease. Logistic Regression gives accuracy of around 91% for prediction of Vector Borne Disease.

KEYWORDS: Vector Borne Diseases, Machine Learning, Logistic Regression, Prediction Model Disease Prevention.

I. INTRODUCTION

Human beings are considered as the top of creation of God in this Universe. Since beginning human beings are continuously endeavoring for improving their life style. Developments in science and technology have made the life easier to live. But still, there exists life threatening challenges to protect human life from attack of various diseases which are also evolving as the life evolves. These diseases affect from person to person, animal to person or from any other environmental factor. With the passage of time, Humans have developed medical diagnosis system to identify and cure these diseases with the help of science and technology. Medical science has invented various medical tests to diagnose different diseases and hence plays an important role to save the life of human kind. With the advent of Artificial Intelligence and Machine Learning techniques, intelligent machines also came into existence which is capable of monitoring and reporting the health parameters and patient's condition continuously in various medical settings. Still the research area of vector borne diseases is in its nascent stages and requires thorough studies in various aspects pertaining to the vector borne diseases such as study of different vectors affecting human being and animals, trapping of different type of vectors, study of environment factors of different geographical locations case study of diagnoses and treatment of these diseases, designing new models for detection and prevention of these diseases at early stages and technologies to prevent outbreak of these diseases at early stages.

Vector borne diseases affect a large part of world population. Every country in the world is suffering from these diseases annually. These diseases have become extensive burden for the developing countries like India, Bangladesh, Nepal etc. Common Vector borne diseases namely Vector Borne Disease, malaria, chikungunya and kala azar are seasonal diseases spreading due to environmental changes and geographical situations.

In India, Vector borne diseases are spreading in different parts of the country. Due to diversity in environment, climate and geographical situations, there is a big challenge for controlling and preventing these kinds of diseases. There are five major vector borne diseases namely Malaria, Vector Borne Disease, Chikungunya, Kala Azar, which are very common diseases in India. There is a big challenge to prevent and control these diseases in India. In rural area, it has become more complex job to control these vector diseases due to the lack of proper medical facilities and infrastructure. So, it becomes necessary to find solution to these outbreaks using latest 2 technologies such as Information Technology and Machine Learning. There is also a dire need of technology based platforms which can reduce the gap between urban and rural by providing them medical facilities with the help of Machine learning. This chapter provides an overview of machine learning and its application on vector borne diseases.

II. LITERATURE SURVEY

V. Janani et. al(2020) Machine learning (ML) is that the application of AI (AI) that has systems the power to be told automatically from experience. The first aim of ML is to permit computers learn without human intervention or

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106023

assistance. It easily identifies trends and patterns, continuous improvement, handling multidimensional and multi-variety data. Vector Borne Disease is that the most virally occurring mosquito-borne disease in recent days. The Multilayer Perception (MLP) algorithm in Machine Learning is employed to realize accuracy in analyzing and predicting Vector Borne Disease. The parameters of the Vector Borne Disease integrated model are identified using an optimization-based methodology in multiple stages. The prediction of Vector Borne Disease using the MLP algorithm is allotted by three phases. The initial phase of Vector Borne Disease prediction is data visualization and pre-processing which is implemented by SMO (Sequential Minimal Optimization). SMO is an algorithm for solving the quadratic programming problem that arises during the training of Support Vector Machines (SVM). The second phase is that the MLP feature selection algorithm which has a leverage backward logistic regression risk analysis. The ultimate phase includes feature reduction by MLP could be a part of dimensionality reduction within the dataset. The implementation of Vector Borne Disease prediction is completed by the WEKA tool. WEKA tool could be a collection of machine learning algorithms for data processing tasks. This tool will be applied to a dataset directly or called from the java code and contains tools for data preprocessing, classification, regression, clustering, association rules, and visualization..

Mahalakshmi and G. Suseendran(2020) The Zika virus presents an extraordinary public health hazard after spreading from Brazil to the Americas. In the absence of credible forecasts of the outbreak's geographic scope and infection frequency, international public health agencies were unable to plan and allocate surveillance resources efficiently. An RNA test will be done on the subjects if they are found to be infected with Zika virus. By training the specified characteristics, the suggested Hybrid Optimization Algorithm such as multilayer perceptron with probabilistic optimization strategy gives forth a greater accuracy rate. The MATLAB program incorporates numerous machine learning algorithms and artificial intelligence methodologies. It reduces forecast time while retaining excellent accuracy. The projected classes are encrypted and sent to patients. The Advanced Encryption Standard (AES) and TRIPLE Data Encryption Standard (TEDS) are combined to make this possible (DES). The experimental outcomes improve the accuracy of patient results communication. Cryptosystem processing acquires minimal timing of 0.15 s with 91.25 percent accuracy.

Singh et al. explored advanced image processing techniques for detecting oral cancer in microscopic biopsy images. Their approach involved transforming images from RGB to Lab color space for better color representation, employing k-means clustering for color classification, and segmenting nuclei to extract relevant features. These features were then classified to distinguish cancerous from non-cancerous lesions. This method highlights the potential of histological analysis and image enhancement for early detection and diagnosis of oral cancer, particularly in regions like the tongue, cheek lining, and palate, where such malignancies commonly occur.

Saranya et al. proposed an innovative DRL-LOA framework that combines Deep Reinforcement Learning (DRL) and Lion Optimization Algorithm (LOA) to enhance decision-making in healthcare cloud-edge networks. Addressing limitations in traffic-centric methods, this approach integrates Graph Convolutional Networks (GCN) to extract network structure data, which is fused with traffic features to create node vectors. These vectors are utilized by a deep Q-network to optimize task allocation at edge nodes, enabling FDNN classifiers to effectively categorize medical data. Simulation results demonstrate significant improvements, including a 35.7% reduction in power consumption and a 27.9% decrease in latency, showcasing the framework's efficacy in optimizing medical data handling and load balancing in cooperative cloud-edge systems.

Jahnavi et al. (2022) explored machine learning algorithms like logistic regression, random forest, decision trees, and SVM for cancer prediction. Their study highlights ML's potential in enhancing diagnostic accuracy, early detection, and personalized treatment. The comparative analysis emphasizes algorithm suitability for specific cancer datasets, addressing critical challenges in cancer evaluation.

Sivaranjani. R and Dr. N.Yuvaraj(2021) Cardiovascular Disease is the silent killer and it is one of the leading cause for global death annually. The percentage of premature death varies from 7% in high-income countries and 43% in low-income countries. It is mainly due to lifestyle changing factors such as obesity, diabetes, etc. While working to reduce earlier deaths, it is revealed, how important the earlier prediction of heart disease is. In the medical field, diagnosing heart disease earlier is a difficult task for medical practitioner since it depends on combining clinical and pathological data. The purpose of this paper is to implement a medical prediction support system for predicting cardiac disease. Deep learning approach based computational model is designed for diagnosis. This proposed system has three main steps. First, a dataset with 13 attributes (13 clinical features) from the website is collected. Second, the datasets are trained using an algorithm called artificial neural network with back propagation technique. It can have one or more hidden layers in order to get higher accuracy. Finally, Cardiac Function Prediction System (CFPS) which is an

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106023

interactive GUI is developed, where the user can enter the clinical features and get to know the current status of a patient's health. This system enhances medical care and reduces the treatment cost. This system will act as a promising tool for the medical practitioner for proper diagnosis

Md. Osman Goni Nayeem et.al(2020) ANN has been proved as a powerful discriminating classifier for tasks in medical diagnosis for early detection of diseases. In our research, ANN has been used for predicting three different diseases (heart disease, liver disorder, lung cancer). Feed-forward back propagation neural network algorithm with Multi-Layer Perception is used as a classifier to distinguish between infected or non-infected person. The results of applying the ANNs methodology to diagnosis of these disease based upon selected symptoms show abilities of the network to learn the patterns corresponding to symptoms of the person. In our proposed work, Multi-Layer Perceptron with having 2 hidden layer is used to predict medical diseases. Here in case of liver disorder prediction patients are classified into four categories: normal condition, abnormal condition (initial), abnormal condition and severe condition. This neural network model shows good performance in predicting disease with less error.

III. EXISTING SYSTEM

- This project introduces a novel approach for Vector Borne Disease fever classification based on online learning paradigms. The proposed approach is suitable for practical implementation as it enables learning using only a few training samples. With time, the proposed approach is capable of learning incrementally from the data collected without need for retraining the model or redeployment of the prediction engine. Additionally, we also provide a comprehensive evaluation of machine learning methods for prediction of Vector Borne Disease fever.
- The input to the proposed pipeline comprises of recorded patient symptoms and diagnostic investigations. Offline classifier models have been employed to obtain baseline scores to establish that the feature set is optimal for classification of Vector Borne Disease. The primary benefit of the online detection model presented in the paper is that it has been established to effectively identify patients with high likelihood of Vector Borne Disease.

Drawbacks

- It does not execute very well when the data set has more sound i.e. target classes are overlapping.
- It doesn't perform well when we have large data set because the required training time is higher.
- A large number of trees can make the algorithm too slow and ineffective for real-time predictions..

IV. PROPOSED SYSTEM

- World Health Organization recognizes that more than 17% of all infectious diseases is accounted by Vector-borne diseases. Around the world, in 97 countries malaria transmission occurs which puts about 3.4 billion people at risk. World's population of over 40% is at risk of Vector Borne Disease. In 2006, the outbreak spread of Chikungunya was in several countries, including India cases were reported. Countries like the United States of America, reported vector-borne disease cases in 2013.The Global Vector Control Response recognize vector control as elementary approach to prevent the vector borne diseases and as a response in order to outbreak.
- This project focused on effective prediction of the vector borne disease outbreak of three diseases (Chikungunya, Malaria, Vector Borne Disease,) across the Indian-subcontinent. We have examined and refined our model over data collected across India in. We have put forward a Logistic Regression outbreak risk prediction algorithm using contrasting data.

V. ADVANTAGES

- It is easier to implement, interpret, and very efficient to train.
- It can easily extend to multiple classes (multinomial regression) and a natural probabilistic view of class predictions.
- It is very fast at classifying unknown records

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106023

VI. METHODOLOGY

Modules

Data Collection

Primary data of 250 Patients of Vector Borne Disease test reports are collected from Tuli Diagnostic Centre, Amritsar in which Vector Borne Disease Test reports includes Ns1 and IgM reports. The symptoms information of each patient is collected telephonically.

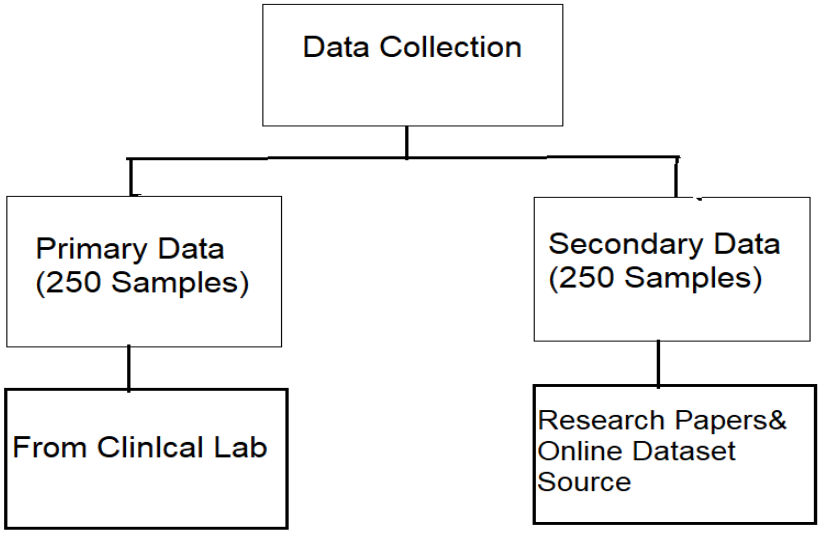


Fig : Data Collection

Raw data Preparations

The first step for developing any prediction model is to prepare a fine and error free dataset. It should have good quality as well as quantity of data effect model performance. It is necessary to covert the unstructured and semi- structured data into the structured form.

Feature selection

Feature selection is an important process to develop a machine learning prediction model. A well designed dataset is the foundation of successful machine learning prediction model. Therefore, it becomes necessary to identify features related to Vector Borne Disease. In this section, the investigator has studied twenty research papers related to Vector Borne Disease symptoms and parameters. From this study, the investigator selects important symptoms as features from each research paper. From here, the investigator prepares a table of symptoms related to Vector Borne Disease diseases. In the table, the researcher has calculated frequency of occurrence of each symptom. This table shows higher frequency symptoms that are found to be significant for detecting the Vector Borne Disease at early stage using machine learning techniques and methods.

Identification of features pertaining to Vector Borne Disease

This section focuses on identifying symptoms and their parameters necessary to identify Vector Borne Disease cases from literature review. For this purpose, the investigator efforts are to identify twenty research papers pertaining to identify various symptoms related to the Vector Borne Disease that is shown in table .

The table shows important factors for detection of Vector Borne Disease. This table also shows frequency of each symptom pertaining to Vector Borne Disease. The important factors pertaining to Vector Borne Disease detection are as shown in figure. In this figure fever, Muscle Pain and Vomiting etc. are shown through bars. An investigator calculates the rank of each symptom as shown in table.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106023

Serial Number	Reference	Fever	Headache	Skin Rash	Joint Pain	Muscle Pain	Pain Behind Eyes	Nausea	Abdominal Pain	Itching	Fatigue	Diarrhoea	Yellow Skin	Chills	Cough	Myalgia	Vomiting	Bleeding from Node and Gum	Rhinorrhoea	Fatigue	Hematemesis	Enlarge Lever	Absence of Cough	Arthralgia	Respiratory	Shock	Bone Pain	Sore Throat	Gastric Bleeding	Swelling	Respiratory System	Petechiae
1	[66]	√	√	√					√							√																
2	[67]	√	√	√						√						√	√	√														
3	[68]	√	√	√					√									√	√	√	√	√	√						√			
4	[69]	√	√	√			√									√		√					√									
5	[70]	√	√	√					√									√														
6	[71]	√	√						√					√			√	√					√									
7	[72]	√	√	√	√	√	√																									
8	[73]	√	√	√			√	√	√		√					√	√	√		√				√	√	√			√		√	
9	[74]	√	√				√	√				√		√	√			√														
10	[75]	√	√		√		√	√	√			√					√	√														
11	[76]	√	√	√				√	√	√		√	√			√	√										√	√				
12	[77]	√		√								√			√		√												√	√		
13	[78]	√	√				√	√						√		√								√								√
14	[79]	√	√		√		√	√								√												√				
15	[80]	√	√	√					√									√														
16	[81]	√	√									√			√	√	√	√						√								√
17	[82]	√	√	√			√		√							√	√															
18	[83]	√	√	√			√	√				√				√	√							√			√					
19	[84]		√	√	√		√	√	√			√		√			√							√				√				
20	[85]	√	√	√					√			√					√											√				
Frequency		19	19	13	4	1	10	8	11	2	11	8	1	19	3	10	10	10	1	2	1	1	3	5	1	1	2	4	3	1	1	2

Fig : Identification of features pertaining to Dengue Disease

Split Dataset Module

This section describes how a large dataset is split into training and test dataset for building prediction model for vector borne diseases. The investigator has split dataset into two parts in the ratio of 70:30. The first part is a training set while the second part is of Test dataset which has shown in figure.

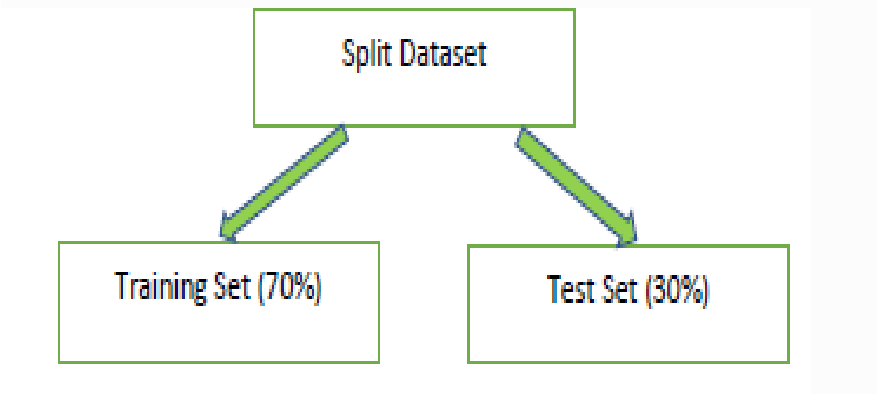


Fig : Ratio of data split in 70:30 ratios of dataset

Model Building Module

This model building module describes the implementation of Logistic Regression Classification machine learning algorithm for the proposed prediction model. The detailed information and implementation of Logistic Regression Classification for the proposed prediction model.

Prediction Module

This section describes the state of art techniques used to make prediction by the proposed prediction model for vector borne Vector Borne Disease. This proposed model is capable of diagnosing two cases of Vector Borne Disease namely: symptoms without clinical reports and symptoms with clinical reports as shown in figure.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106023

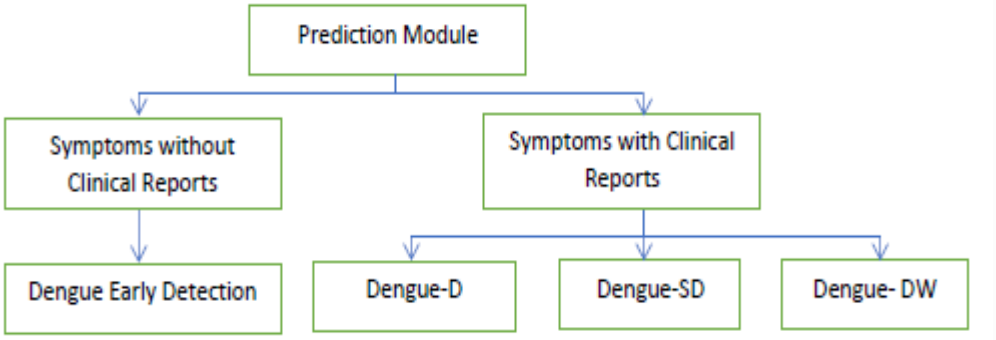


Fig : Prediction module of proposed prediction model for vector borne diseases (Dengue)

VII. SYSTEM ARCHITECTURE

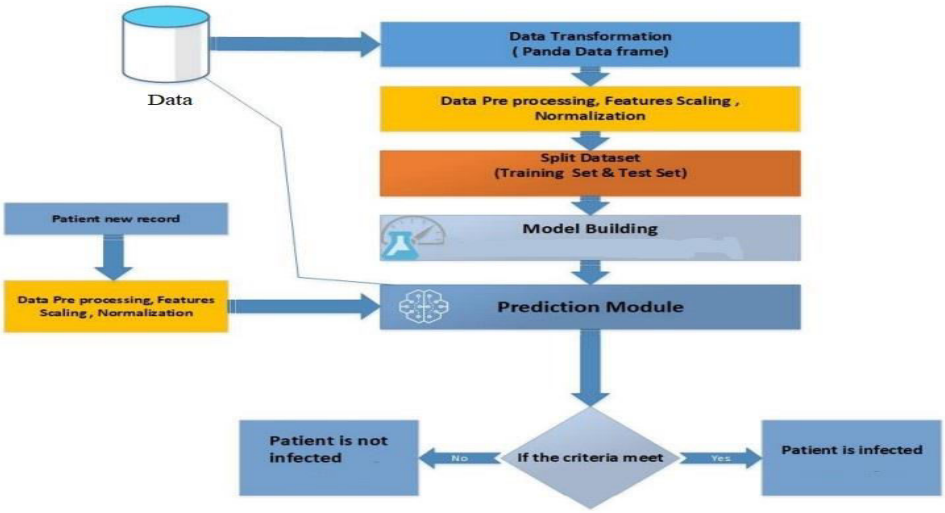


Fig : System Architecture

Explanation:

- **Data Transformation:** Load the data into a Pandas Data Frame**.**
- **Data Preprocessing:** Handle missing values, encode categorical variables, and prepare the data for modeling**.**
- **Feature Scaling and Normalization**: Scale features to ensure they are on a similar scale.
- **Split Dataset**: Divide the dataset into training and testing sets**.**
- **Model Building:** Train a machine learning model.
- **Prediction Module:** Process new patient records and make predictions.

VIII. IMPLEMENTATION

Logistic regression is a statistical method used for predicting the probability of a binary outcome based on one or more independent variables. It is a classification algorithm often applied in machine learning and statistics.

1. Binary Outcome: Logistic regression predicts categorical outcomes, usually binary (e.g., yes/no, true/false, 0/1).
2. Logit Function: It uses the logistic (or sigmoid) function to map predicted values to probabilities between 0 and 1.
3. Linear Relationship: Logistic regression assumes a linear relationship between the independent variables and the log-odds of the dependent variable.

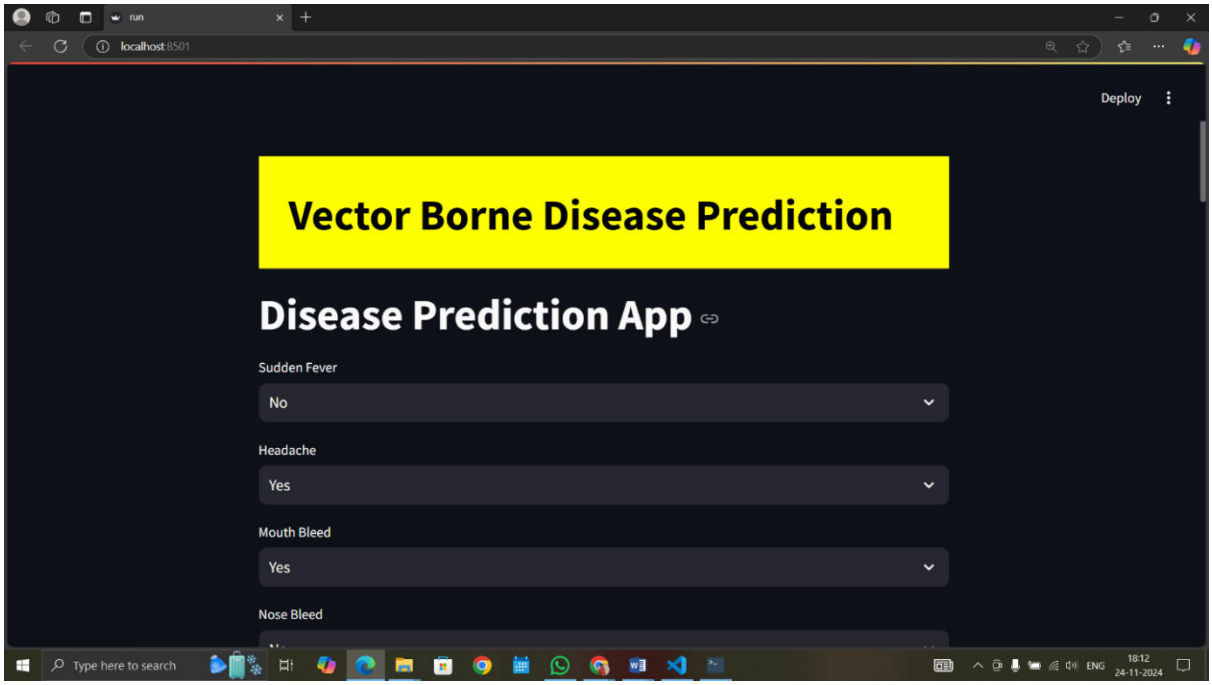
International Journal of Advanced Research in Education and TechnologY(IJARETY)

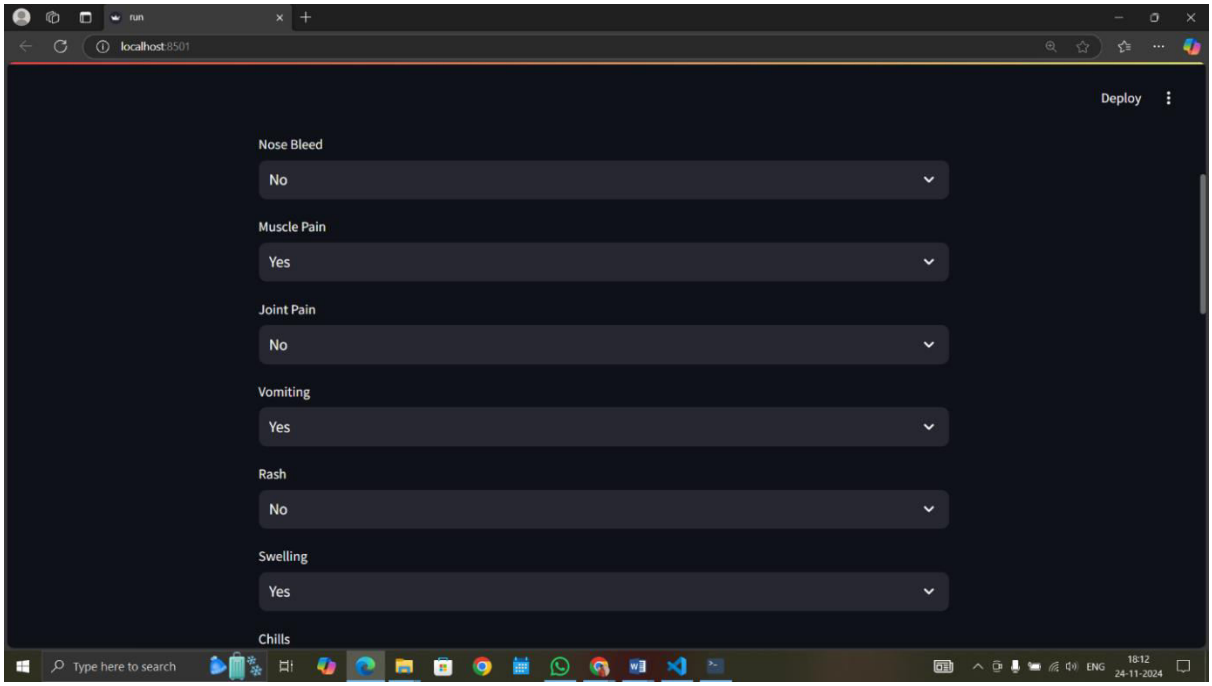
| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106023

IX. EXPERIMENTAL RESULTS



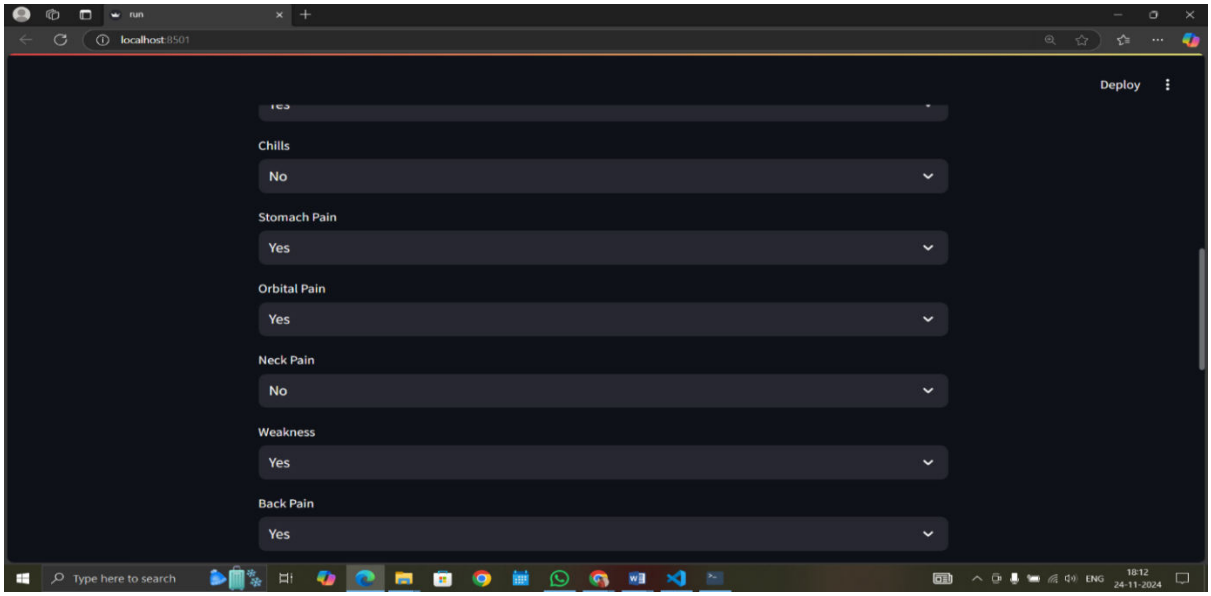


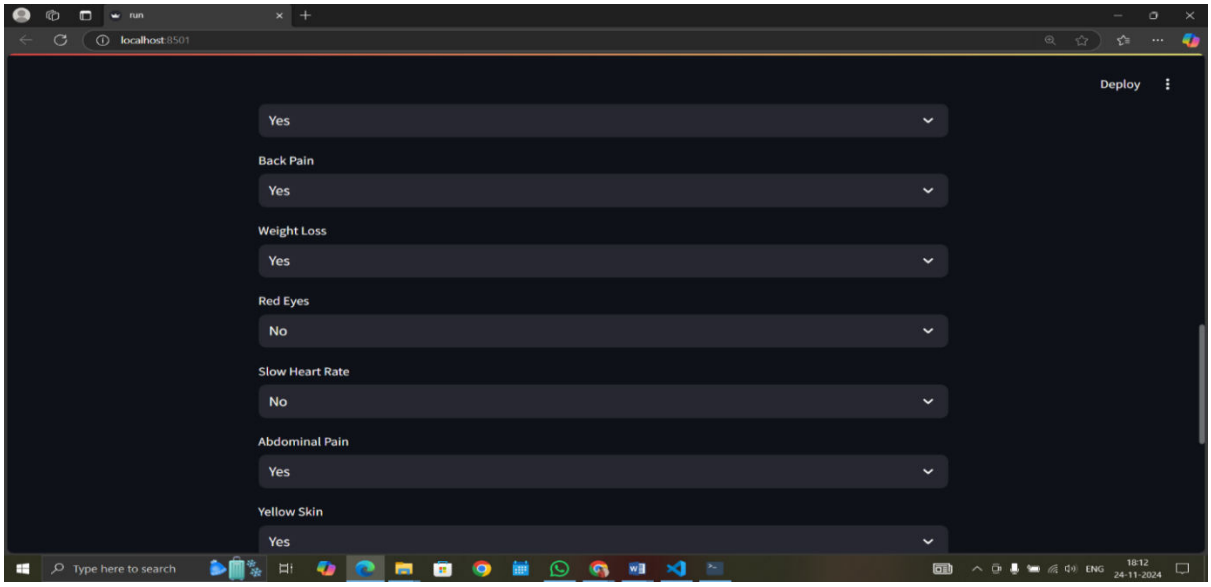
International Journal of Advanced Research in Education and TechnologY(IJARETY)

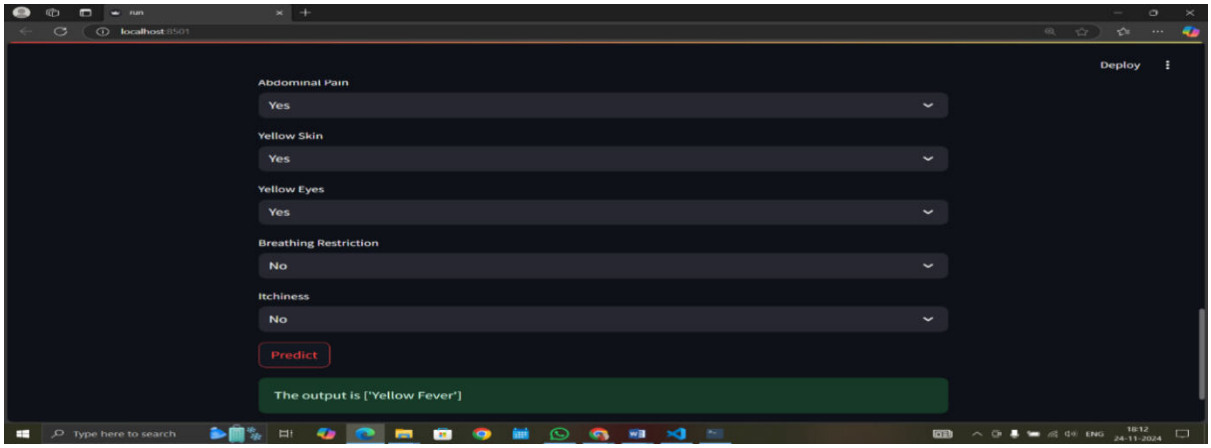
| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106023







Figures: Outputs of Vector Borne Diseases

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 6, November-December 2024 ||

DOI:10.15680/IJARETY.2024.1106023

The user need to enter the symptoms which he has been suffering, by selecting the options in form.

REFERENCES

1. Dr. K. Usha Rani on "Analysis of heart diseases dataset using neural network approach" International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.1, No.5, September 2011
2. Shraddha Shiv hare and Rajesh Shrivastava on "Automatic Bone Marrow White Blood Cell classification using Morphological Granulometric feature of Nucleus" Oriental Journal Of Computer Science & Technology Vol. 5, No. (1): Pgs. 75-85, June 2012
3. Md. Osman Goni Nayeem, Maung Ning Wan and Md. Kamrul Hasan on "Prediction of disease level using multilayer perceptron of Artificial Neural Network for patient monitoring" International Journal of Soft Computing and Engineering (IJSCE) Volume-5 Issue-4: Pgs. 17-23, September 2015
4. Sivaranjani.R and Dr.N.Yuvaraj on "Artificial Intelligence model for earlier prediction of cardiac functionalities using multilayer perceptron" International Conference on Physics and Photonics Processes in Nano Sciences
5. Marios Anthimopoulos, Stergios Christodoulidis, Lukas Ebner, Andreas Christe, and Stavroula Mougiakakou on "Lung pattern classification for interstitial lung diseases using a deep Convolutional Neural Network" IEEE Transactions On Medical Imaging, VOL. 35, NO. 5, MAY 2016
6. Rahib H. Abiyev and Mohammad Khaleel Sallam Maaitah "Deep Convolutional Neural Networks for chest diseases detection" Journal of Healthcare Engineering Volume 2018
7. Agus Qomaruddin Munir and Edi Winarko on "Classification model disease risk areas endemicity Vector Borne Disease Fever outbreak based prediction of Patients, death, IR and CFR using forecasting techniques" International Journal of Computer Applications, March 2015
8. Juan M. Scavuzzoa , Francisco Truccoa , Manuel Espinosac , Carolina B. Taurob , Marcelo Abrilc , Carlos M. Scavuzzob ,and Alejandro C. Freryd Modeling on "Vector Borne Disease vector population using remotely sensed data and Machine Learning" in press.
9. Mahalakshmi and G. Suseendran on "Prediction of Zika Virus by multilayer perceptron neural network (MLPNN) using Cloud B" International Journal of Recent Technology and Engineering (IJRTE), Volume-8 Issue-2S11: Pgs. 249-254, September 2019
10. V. Janani , N. Maadhuryaa , D. Pavithra and S. Ramya Sree on "Vector Borne Disease Prediction using multilayer perceptron - a Machine Learning approach" International Journal of Research in Engineering, Science and Management Volume-3, Issue-3, March 2020
11. Oladimeji Mudele ,Fábio M. Bayer, Lucas Zanandrez, Alvaro E. Eiras and Paolo Gamba on "Modeling the temporal population distribution of Ae. Aegypti mosquito using big Earth observation data" in press.
12. Vijeta Sharma ,Ajai Kumar ,Lakshmi Panat ,Dr. Ganesh Karajkhede and Anuradha lele on "Malaria outbreak prediction model using Machine Learning" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 4 Issue 12, December 2015
13. Jahnavi, Y., Kumar, P. N., Anusha, P., & Prasad, M. S. (2022, November). Prediction and Evaluation of Cancer Using Machine Learning Techniques. In International Conference on Sustainable and Innovative Solutions for Current Challenges in Engineering & Technology (pp. 399-405). Singapore: Springer Nature Singapore.
14. Saranya, S. S., Anusha, P., Chandragandhi, S., Kishore, O. K., Kumar, N. P., & Srihari, K. (2024). Enhanced decision-making in healthcare cloud-edge networks using deep reinforcement and lion optimization algorithm. Biomedical Signal Processing and Control, 92, 105963.
15. Singh, A., Kumar, T. C. A., Mithun, T., Majji, S., Rajesh, M., & Anusha, P. (2021, December). Image Processing Approaches for Oral Cancer Detection in Color Images. In 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 817-821). IEEE.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 7.394

www.ijarety.in editor.ijarety@gmail.com